

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 10, 2003, 21:32:07 ; Search time 8449.13 Seconds
(without alignments)
117.112 Million cell updates/sec

Title: US-09-913-524-33

Perfect score: 34

Sequence: 1 aggcctccggaggaaacnctgccatgccaact 34

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl :

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vt.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_pi.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vt.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rdt.*
- 36: em_htg_nam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. Is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 33 | 97.1 | 847 | 9 | HSINHAG2 |
| 2 | 33 | 97.1 | 1134 | 6 | A14420 |
| 3 | 33 | 97.1 | 1237 | 6 | I01843 |
| 4 | 33 | 97.1 | 1237 | 6 | I01843 |
| 5 | 33 | 97.1 | 1237 | 6 | I01843 |
| 6 | 33 | 97.1 | 1237 | 6 | I01843 |
| 7 | 33 | 97.1 | 1237 | 6 | I01843 |
| 8 | 33 | 97.1 | 1237 | 6 | I01843 |
| 9 | 33 | 97.1 | 1237 | 6 | I01843 |
| 10 | 33 | 97.1 | 1237 | 6 | I01843 |
| 11 | 33 | 97.1 | 1237 | 6 | I01843 |
| 12 | 33 | 97.1 | 1237 | 6 | I01843 |
| 13 | 33 | 97.1 | 1237 | 6 | I01843 |
| 14 | 33 | 97.1 | 1237 | 6 | I01843 |
| 15 | 33 | 97.1 | 1237 | 6 | I01843 |
| 16 | 26.6 | 78.2 | 994 | 4 | ECU21219 |
| 17 | 26.6 | 78.2 | 1144 | 10 | RATINHAB2 |
| 18 | 26.6 | 78.2 | 1286 | 4 | HRSIASP |
| 19 | 26.6 | 78.2 | 1372 | 10 | AF432351 |
| 20 | 26.6 | 78.2 | 1561 | 10 | RATINHA |
| 21 | 26.6 | 78.2 | 101356 | 2 | AC112361 |
| 22 | 25 | 73.5 | 857 | 4 | SHPINHA |
| 23 | 25 | 73.5 | 1160 | 10 | MMINAS |
| 24 | 25 | 73.5 | 1182 | 4 | BOVINHA |
| 25 | 25 | 73.5 | 1183 | 6 | A14416 |
| 26 | 25 | 73.5 | 1183 | 10 | MUSINHIB02 |
| 27 | 25 | 73.5 | 1328 | 10 | MMALPHA |
| 28 | 23.4 | 68.8 | 731 | 4 | AY028465 |
| 29 | 23.4 | 68.8 | 731 | 4 | AY028466 |
| 30 | 23.4 | 68.8 | 1272 | 4 | PIGINHA |
| 31 | 23.4 | 68.8 | 1333 | 4 | SSINHAR |
| 32 | 23.4 | 68.8 | 1343 | 6 | I01835 |
| 33 | 23.4 | 68.8 | 1343 | 6 | I05231 |
| 34 | 23.4 | 68.8 | 1343 | 6 | I05231 |
| 35 | 23.4 | 68.8 | 1343 | 6 | I05231 |
| 36 | 23.4 | 68.8 | 1343 | 6 | I05231 |
| 37 | 22 | 64.7 | 237982 | 2 | AC093483 |
| 38 | 21.6 | 63.5 | 180525 | 2 | AC104750 |
| 39 | 21.6 | 63.5 | 187740 | 2 | AC126863 |
| 40 | 21 | 61.8 | 162328 | 9 | AP000350 |
| 41 | 21 | 61.8 | 236809 | 2 | AC103535 |
| 42 | 20.8 | 61.2 | 5108 | 9 | HSN802883 |
| 43 | 20.8 | 61.2 | 5145 | 6 | AX127534 |
| 44 | 20.8 | 61.2 | 5145 | 9 | AB037848 |
| 45 | 20.8 | 61.2 | 11158 | 1 | AE004682 |

ALIGNMENTS

| | | | | | |
|------------|---|--------|-----|--------|-----------------|
| RESULT 1 | HSINHAG2 | 847 bp | DNA | linear | PRI 10-FEB-1997 |
| LOCUS | H.sapiens irhA gene, exon 2. | | | | |
| DEFINITION | X04446 | | | | |
| ACCESSION | X04446.1 | | | | |
| VERSION | GI:33924 | | | | |
| KEYWORDS | glycoprotein; glycoprotein hormone; hormone; inhibin; preprohormone. | | | | |
| SOURCE | Homo sapiens | | | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. | | | | |
| REFERENCE | 1 (bases 1 to 847) | | | | |
| AUTHORS | Stewart,A.G., Milborrow,H.M., Ring,J.M., Crowther,C.E. and Forage,R.G. | | | | |

| | | | | | | | | | |
|--|-----|-----------------------------------|-----|--|--|--|--|--|--|
| Best Local Similarity 97.1%; Pred. No. 0.0097; | | | | | | | | | |
| Matches 33; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | | | | | | | | | |
| | | | | | | | | | |
| QY | 1 | AGGCTCCGGAGAACCGNCTGCCCATGCCCAACT | 34 | | | | | | |
| Db | 758 | AGGCTCCGGAGAACCGGCTGCCCATGCCCAACT | 791 | | | | | | |
| | | | | | | | | | |
| RESULT 3 | | | | | | | | | |
| I01843 | | | | | | | | | |
| LOCUS I01843 1237 bp ss-DNA linear PAT 21-MAY-1993 | | | | | | | | | |
| DEFINITION Sequence 10 from Patent US 4798885. | | | | | | | | | |
| ACCESSION I01843 | | | | | | | | | |
| VERSION I01843.1 GI:269775 | | | | | | | | | |
| KEYWORDS | | | | | | | | | |
| SOURCE Unknown. | | | | | | | | | |
| ORGANISM Unknown. | | | | | | | | | |
| REFERENCE 1 (bases 1 to 1237) | | | | | | | | | |
| AUTHORS Mason,A.J. and Seeburg,P.H. | | | | | | | | | |
| TITLE Compositions of hormonally active human and porcine inhibin containing an alpha chain and 62 chain. | | | | | | | | | |
| JOURNAL Patent: US 4798885-A 10 17-JAN-1989; | | | | | | | | | |
| Genentech, Inc.; South San Francisco, CA | | | | | | | | | |
| FEATURES | | | | | | | | | |
| source Location/Qualifiers | | | | | | | | | |
| 1..1237 | | | | | | | | | |
| /organism="unknown" | | | | | | | | | |
| BASE COUNT 210 a 431 c 346 g 250 t | | | | | | | | | |
| ORIGIN | | | | | | | | | |
| Query Match 97.1%; Score 33; DB 6; Length 1237; | | | | | | | | | |
| Best Local Similarity 97.1%; Pred. No. 0.0096; | | | | | | | | | |
| Matches 33; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | | | | | | | | | |
| | | | | | | | | | |
| QY | 1 | AGGCTCCGGAGAACCGNCTGCCCATGCCCAACT | 34 | | | | | | |
| Db | 708 | AGGCTCCGGAGAACCGGCTGCCCATGCCCAACT | 741 | | | | | | |
| | | | | | | | | | |
| RESULT 4 | | | | | | | | | |
| I05264 | | | | | | | | | |
| LOCUS I05264 1237 bp DNA linear PAT 02-DEC-1994 | | | | | | | | | |
| DEFINITION Sequence 21 from Patent EP 0222491. | | | | | | | | | |
| ACCESSION I05264 | | | | | | | | | |
| VERSION I05264.1 GI:591394 | | | | | | | | | |
| KEYWORDS | | | | | | | | | |
| SOURCE Unknown. | | | | | | | | | |
| ORGANISM Unknown. | | | | | | | | | |
| REFERENCE 1 (bases 1 to 1247) | | | | | | | | | |
| AUTHORS Mason,A.J. and Seeburg,P.H. | | | | | | | | | |
| TITLE Nucleic acid encoding the alpha or beta chains of inhibin and method for synthesizing polypeptides using such nucleic acid | | | | | | | | | |
| JOURNAL Patent: EP 0222491-A1 21 20-MAY-1987; | | | | | | | | | |
| Genentech, Inc.; South San Francisco, CA | | | | | | | | | |
| FEATURES | | | | | | | | | |
| source Location/Qualifiers | | | | | | | | | |
| 1..1237 | | | | | | | | | |
| /organism="unknown" | | | | | | | | | |
| BASE COUNT 210 a 431 c 346 g 250 t | | | | | | | | | |
| ORIGIN | | | | | | | | | |
| Query Match 97.1%; Score 33; DB 6; Length 1237; | | | | | | | | | |
| Best Local Similarity 97.1%; Pred. No. 0.0096; | | | | | | | | | |
| Matches 33; Conservative 0; Mismatches 1; Indels 0; Gaps 0; | | | | | | | | | |
| | | | | | | | | | |
| QY | 1 | AGGCTCCGGAGAACCGNCTGCCCATGCCCAACT | 34 | | | | | | |
| Db | 708 | AGGCTCCGGAGAACCGGCTGCCCATGCCCAACT | 741 | | | | | | |
| | | | | | | | | | |
| RESULT 5 | | | | | | | | | |
| I21913 | | | | | | | | | |
| LOCUS I21913 1237 bp DNA linear PAT 07-OCT-1996 | | | | | | | | | |
| DEFINITION Sequence 40 from patent US 5525488 | | | | | | | | | |

```

ACCESSION      I21913
VERSION        I21913.1  GI:1602267
KEYWORDS
SOURCE        Unknown.
ORGANISM      Unclassified.
REFERENCE      1 (bases 1 to 1237)
AUTHORS      Mason,A.J. and Seeburg,P.H.
TITLE        Nucleic acid encoding the mature .alpha. chain of inhibin and
              method for synthesizing polypeptides using such nucleic acid
JOURNAL       Patent: US 5525488-A 40 11-JUN-1996;
              Location/Qualifiers
FEATURES      source
              1..1237
              /organism="unknown"
BASE COUNT    210 a 431 c 346 g 250 t
ORIGIN
Query Match   97.1%; Score 33; DB 6; Length 1237;
Best Local Similarity 97.1%; Pred. No. 0.0096;
Matches 33; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCTCCGGAGGAGACCGNCTGCCATGCCAACT 34
|||||
Db 708 AGGCTCCGGAGGAGAACCGGCTGCCATGCCAACT 741

RESULT 6
LOCUS         164619
DEFINITION   Sequence 40 from patent US 5665568.
ACCESSION    164619
VERSION      164619.1  GI:2481513
KEYWORDS
SOURCE       Unknown.
ORGANISM     Unclassified.
REFERENCE    1 (bases 1 to 1237)
AUTHORS      Mason,A.J. and Seeburg,P.H.
TITLE        Nucleic acid encoding the mature .beta. sub.A chain of inhibin and
              method for synthesizing polypeptides using such nucleic acid
JOURNAL       Patent: US 5665568-A 40 09-SEP-1997;
              Location/Qualifiers
FEATURES      source
              1..1237
              /organism="unknown"
BASE COUNT    210 a 431 c 346 g 250 t
ORIGIN
Query Match   97.1%; Score 33; DB 6; Length 1237;
Best Local Similarity 97.1%; Pred. No. 0.0096;
Matches 33; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCTCCGGAGGAGACCGNCTGCCATGCCAACT 34
|||||
Db 708 AGGCTCCGGAGGAGAACCGGCTGCCATGCCAACT 741

RESULT 7
LOCUS         187901
DEFINITION   Sequence 40 from patent US 5716810.
ACCESSION    187901
VERSION      187901.1  GI:3407841
KEYWORDS
SOURCE       Unknown.
ORGANISM     Unclassified.
REFERENCE    1 (bases 1 to 1237)
AUTHORS      Mason,A.J. and Seeburg,P.H.
TITLE        Nucleic acid encoding the mature .beta. sub.B chain of inhibin and
              method for synthesizing polypeptides using such nucleic acid
JOURNAL       Patent: US 5716810-A 40 10-FEB-1998;
              Location/Qualifiers
FEATURES      source
              1..1237

```

```

BASE COUNT    210 a 431 c 346 g 250 t
ORIGIN
Query Match   97.1%; Score 33; DB 6; Length 1237;
Best Local Similarity 97.1%; Pred. No. 0.0096;
Matches 33; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCTCCGGAGGAGACCGNCTGCCATGCCAACT 34
|||||
Db 708 AGGCTCCGGAGGAGAACCGGCTGCCATGCCAACT 741

RESULT 8
LOCUS         HUMINHAA
DEFINITION   Human ovarian alpha-inhibin mRNA.
ACCESSION    M13144
VERSION      M13144.1  GI:186412
KEYWORDS
SOURCE       Human polycystic ovarian cDNA to mRNA, clones lambda-hin-alpha-
              [2.6].
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 1237)
AUTHORS      Mason,A.J., Niall,H.B. and Seeburg,P.H.
TITLE        Structure of two human ovarian inhibins
JOURNAL       Biochem. Biophys. Res. Commun. 135 (3), 957-964 (1986)
MEDLINE      86186863
PUBMED       3754442
FEATURES      Location/Qualifiers
              1..1237
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /map="4q33-qter"
              1..1237
              /gene="INH"
              <1..1058
              /gene="INH"
              /note="inhibin A"
              /codon_start=3
              /protein_id="AAA59167.1"
              /db_xref="GI:186413"
              /db_xref="GBB:G00-120-100"
              /translation="GVSSQGLELAKRLVLAQVRLALDLALGPVAVTREGDQGVRRLLP
              RRHALGCTHRTGSEFEFEEDVSOAIFVATDASCDKSAACLAQAEAGLIFYMRRP
              SQHRSKQVTSQILWHITGLDRQGTAAANSSEPLGLLALSGGPVAVPMSLGHAPPH
              WAWHIAISGLITHTPVLVLGKPTCTCSARPEATPFLVAHTRTPPSGGERARRS
              TPERSWPSFSAIRLLQRPPEPAHANCHRVALNFSQELGWRWIVPPSTFIHYC
              HGGGLHLPNLSLHVPGAPPTPAQPSLIPGAQPCCAALPGTMRPLHVRTTSDGYS
              FRYETVPLNLTGHCACI"
BASE COUNT    210 a 431 c 346 g 250 t
ORIGIN
Query Match   97.1%; Score 33; DB 9; Length 1237;
Best Local Similarity 97.1%; Pred. No. 0.0096;
Matches 33; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCTCCGGAGGAGACCGNCTGCCATGCCAACT 34
|||||
Db 708 AGGCTCCGGAGGAGAACCGGCTGCCATGCCAACT 741

RESULT 9
LOCUS         IO0529
DEFINITION   Sequence 1 from Patent US 4737578.
ACCESSION    IO0529
VERSION      IO0529.1  GI:268923
KEYWORDS
SOURCE       Unknown.
ORGANISM     Unknown.

```

Unclassified.

REFERENCE 1 (bases 1 to 1338)
 AUTHORS Evans,K.M., Rosenfeld,M.G., Cerelli,G., Mayo,K.E., Spiess,J.,
 Rivier,J.E.F. and Vale,W.W. Jr.
 TITLE Human inhibin
 JOURNAL Patent: US 4737578-A 1 12-APR-1988;
 The Salk Institute for Biological Studies; San Diego, CA

FEATURES
 source 1..1338
 location/Qualifiers
 BASE COUNT 232 a 433 c 417 g 256 t
 ORIGIN

Query Match 97.1%; Score 33; DB 6; Length 1338;
 Best local Similarity 97.1%; Pred. No. 0.0096;
 Matches 33; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGCTCCGGAGAACGCTGCCATGCCCAACT 34
 |||||
 Db 895 AGGCTCCGGAGAACGCTGCCATGCCCAACT 928
 |||||

RESULT 10
 HUMINHA
 LOCUS Human inhibin A-subunit mRNA, complete cds.
 DEFINITION M13981
 ACCESSION M13981.1 GI:186410
 VERSION M13981.1
 KEYWORDS inhibin.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1338)
 AUTHORS Evans,K.E., Cerelli,G.M., Spiess,J., Rivier,J., Rosenfeld,M.G.,
 Mayo,K.E., and Vale,W.
 TITLE Inhibin A-subunit cDNAs from porcine ovary and human placenta
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 83 (16), 5849-5853 (1986)
 MEDLINE 86287450
 PUBMED 3016724

FEATURES
 source 1..1338
 location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="2q33-qter"
 1..1338
 /gene="INHA"
 145..1245
 /gene="INHA"
 /note="inhibin A-subunit precursor"
 /codon_start=1
 /protein_id="AAA59166.1"
 /db_xref="GI:307068"
 /db_xref="DB:GDB:600-120-100"
 /translation="MVLHLLFLLLTPGGHSCQGLELAKVRLAKVRLFDALGPP
 AVREGGDPVRLPRHALGFTHRGSEPEEDYSOAILFPATDASCDKSAARGL
 AQAEGLFRYMPRSOHTRSKQVTSQILWFHGLDROCTAASNSPEPLGLALSFG
 GPVAPNSLGHAPPWAVLHATSLSLLTHPVLVLLRCPLCTCSARPEATPELVAH
 TRTPPSGGERARRSTPLMSWPSPSALRIQLRPEEPAHANCHRVAINISFQELGM
 ERWIVYPPSFIHYHGGCGLLHPNLSLPVPGAPPTPAQPSYLLPQAQPCCAALPGT
 MRPLHVRTISDGGYSEKYEYVFNLLTQHCAC1"
 145..198
 /gene="INHA"
 /note="inhibin A-subunit signal peptide"
 841..1242
 /gene="INHA"
 /product="inhibin A-subunit"
 /db_xref="GI:13623557"
 BASE COUNT 232 a 433 c 417 g 256 t
 ORIGIN 197 bp upstream of PvuII site.

Query Match 97.1%; Score 33; DB 9; Length 1338;
 Best local Similarity 97.1%; Pred. No. 0.0096;
 Matches 33; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGCTCCGGAGAACGCTGCCATGCCCAACT 34
 |||||
 Db 895 AGGCTCCGGAGAACGCTGCCATGCCCAACT 928
 |||||

RESULT 11
 BC006391
 LOCUS Homo sapiens, inhibin, alpha, clone MCC:12547 IMAGE:4126990, mRNA,
 DEFINITION complete cds.
 ACCESSION BC006391
 VERSION BC006391.1 GI:13623556
 KEYWORDS MCC.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1394)
 AUTHORS Strausberg,K.
 TITLE Direct Submission
 JOURNAL Submitted (09-APR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
 COMMENT Contact: MCC help desk
 Email: cga@bs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 DNA Library Arrayed by: The J.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland.
 Web site: http://www.nisc.nih.gov/
 Contact: nisc.mgc@nih.gov
 Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
 Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
 Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R.,
 Lim,M., Maduro,Q.L., Mastello,C., Mastrian,S.D., McCloskey,J.C.,
 McDowell,J., Pearson,R., Snyder,B., Stantropop,S., Thomas,P.J.,
 Tongson,E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
 Zhang,L.-H. and Green,E.D.

Clone distribution: MCC clone distribution information can be found
 through the J.M.A.G.E. Consortium/ILNI at: http://image.llnl.gov
 Series: IRAL Plate: 17 Row: h Column: 17
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 9257223.
 location/Qualifiers
 1..1394
 /organism="Homo sapiens"
 /db_xref="LOCUSID:3623"
 /db_xref="taxon:9606"
 /clone="MGC:12547 IMAGE:4126990"
 /tissue_type="Brain, neuroblastoma"
 /clone_lib="NIH-MGC_19"
 /lab_host="DH10B-R"
 /note="vector: pOTB7"
 90..1190
 /codon_start=1
 /product="inhibin, alpha"
 /protein_id="AAH06391.1"
 /db_xref="GI:13623557"
 /translation="MVLHLLFLLLTPGGHSCQGLELAKVRLAKVRLFDALGPP
 AVREGGDPVRLPRHALGFTHRGSEPEEDYSOAILFPATDASCDKSAARGL
 AQAEGLFRYMPRSOHTRSKQVTSQILWFHGLDROCTAASNSPEPLGLALSFG
 GPVAPNSLGHAPPWAVLHATSLSLLTHPVLVLLRCPLCTCSARPEATPELVAH
 TRTPPSGGERARRSTPLMSWPSPSALRIQLRPEEPAHANCHRVAINISFQELGM
 ERWIVYPPSFIHYHGGCGLLHPNLSLPVPGAPPTPAQPSYLLPQAQPCCAALPGT
 MRPLHVRTISDGGYSEKYEYVFNLLTQHCAC1"
 258 a 469 c 394 g 272 t

BASE COUNT 258 a 469 c 394 g 272 t
 ORIGIN

* 35684 35783: gap of 100 bp
* 35784 36555: contig of 772 bp in length
* 36556 36655: gap of 100 bp
* 36656 37524: contig of 869 bp in length
* 37525 37624: gap of 100 bp
* 37625 38433: contig of 809 bp in length
* 38434 38533: gap of 100 bp
* 38534 39427: contig of 894 bp in length
* 39428 39527: gap of 100 bp
* 39528 40330: contig of 803 bp in length
* 40331 40430: gap of 100 bp
* 40431 41249: contig of 819 bp in length
* 41250 41349: gap of 100 bp
* 41350 42152: contig of 803 bp in length
* 42153 42252: gap of 100 bp
* 42253 43089: contig of 837 bp in length
* 43090 43189: gap of 100 bp
* 43190 43995: contig of 806 bp in length
* 43996 44095: gap of 100 bp
* 44096 44894: contig of 799 bp in length
* 44895 44994: gap of 100 bp
* 44995 45842: contig of 848 bp in length
* 45843 45942: gap of 100 bp
* 45943 46760: contig of 818 bp in length
* 46761 46860: gap of 100 bp
* 46861 47697: contig of 837 bp in length
* 47698 47797: gap of 100 bp
* 47798 48596: contig of 799 bp in length
* 48597 48696: gap of 100 bp
* 48697 49501: contig of 805 bp in length
* 49502 49601: gap of 100 bp
* 49602 50492: contig of 891 bp in length
* 50493 50592: gap of 100 bp
* 50593 51425: contig of 833 bp in length
* 51426 51525: gap of 100 bp
* 51526 52311: contig of 786 bp in length
* 52312 52411: gap of 100 bp
* 52412 53239: contig of 828 bp in length.

FEATURES
source
1. 53239
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-1502"
/clone_lib="RP11-11 Human Male BAC"

HASH COUNT 9751 a 10973 c 14398 g 10550 t 7567 others
ORIGIN

Query Match 97.1%; Score 33; DB 2; Length 53239;
Best Local Similarity 97.1%; Pred. No. 0.0079;
Matches 33; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGCTCCGGAGAACGNCNCTGCGCCATGCCCAACT 34
|||||
Db 1158 AGGCTCCGGAGAACGNCNCTGCGCCATGCCCAACT 1191

RESULT 13
AC009955
LOCUS AC009955 135033 bp DNA linear PRI 08-NOV-2000
DEFINITION Homo sapiens BAC clone RP11-256123 from 2, complete sequence.
ACCESSION AC009955
VERSION AC009955.4 GI:9581957
KEYWORDS HTG.

SOURCE
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 135033)

REFERENCE
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL. Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074

REFERENCE
AUTHORS 2 (bases 1 to 135033)
TITLE Du, H., Maupin, R., Hawkins, M. and Hodges, J.
JOURNAL The sequence of Homo sapiens BAC clone RP11-256123
REFERENCE Unpublished
AUTHORS 3 (bases 1 to 135033)
TITLE Waterston, R.H.
JOURNAL Direct Submission
AUTHORS Submitted (08-SEP-1999) Genome Sequencing Center, Washington
TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63108, USA
AUTHORS 4 (bases 1 to 135033)
TITLE Waterston, R.H.
JOURNAL Direct Submission
AUTHORS Submitted (29-JUL-2000) Genome Sequencing Center, Washington
TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63108, USA
AUTHORS 5 (bases 1 to 135033)
TITLE Waterston, R.H.
JOURNAL Direct Submission
AUTHORS Submitted (10-SEP-2000) Genome Sequencing Center, Washington
TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63108, USA
AUTHORS 6 (bases 1 to 135033)
TITLE Waterston, R.
JOURNAL Direct Submission
AUTHORS Submitted (14-OCT-2000) Department of Genetics, Washington
TITLE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL Waterston, R.
AUTHORS 7 (bases 1 to 135033)
TITLE Direct Submission
AUTHORS Submitted (08-NOV-2000) Department of Genetics, Washington
TITLE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL On Jul 29, 2000 this sequence version replaced gi:8569874.
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0256123

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., paired quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenio, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-51207, 200 base pair

overlap. Actual start of this clone is at base position 1 of
RP11-256123; actual end is at base position 52326 of RP11-51207.

| FEATURES | | Location/Qualifiers | | Source | |
|---------------|--------------|--|--------------------------|--------------|---|
| repeat_region | 1. .135033 | /location="RP11-256123" | /organism="Homo sapiens" | misc_feature | /note="similar to EST AA935178 (NID:g3092335) oh33d03.s1 |
| repeat_region | 5. .136 | /clone_lib="RPC1-11" | /db_xref="taxon:9606" | misc_feature | /note="similar to EST A1012498 (NID:g3226330)" |
| repeat_region | 530. .629 | /rpt_family="MIR" | /map="2" | misc_feature | /note="similar to EST AW918628 (NID:g8084409)" |
| repeat_region | 1324. .1955 | /rpt_family="L2" | /clone="RP11-256123" | misc_feature | /note="similar to EST AW553790 (NID:g7199213)" |
| repeat_region | 1956. .2276 | /rpt_family="Alu" | /clone_lib="RPC1-11" | misc_feature | /note="similar to EST AA475614 (NID:g2203465) vh22a10.r1" |
| repeat_region | 2277. .2490 | /rpt_family="L2" | 5. .136 | misc_feature | /note="similar to EST AA365758 (NID:g2018306)" |
| repeat_region | 4032. .4151 | /rpt_family="MIR" | /rpt_family="MIR" | misc_feature | /note="similar to EST D78947 (NID:g1180820)" |
| repeat_region | 4280. .4634 | /rpt_family="L2" | /rpt_family="L2" | misc_feature | /note="similar to EST AA302774 (NID:g1955159)" |
| repeat_region | 4640. .4949 | /rpt_family="Alu" | /rpt_family="L2" | misc_feature | /note="similar to EST AW451283 (NID:g6992059)" |
| repeat_region | 5165. .5587 | /rpt_family="ERV1" | /rpt_family="L2" | misc_feature | /note="similar to EST AW205277 (NID:g6504681)" |
| misc_feature | 6055. .6213 | /note="similar to EST AW427638 (NID:g6955585)" | /rpt_family="Alu" | misc_feature | /note="similar to EST AA984525 (NID:g3163050) am88c02.s1" |
| misc_feature | 6059. .6213 | /note="similar to EST AW669129 (NID:g7525643)" | /rpt_family="L2" | misc_feature | /note="similar to EST C18330 (NID:g1579932)" |
| repeat_region | 6650. .6960 | /rpt_family="Alu" | /rpt_family="L2" | misc_feature | /note="similar to EST W28969 (NID:g1308917)" |
| repeat_region | 7250. .7330 | /rpt_family="L2" | /rpt_family="L2" | misc_feature | /note="similar to EST W67340 (NID:g1376209) zd43d07.r1" |
| misc_feature | 7955. .8056 | /note="similar to EST AW427638 (NID:g6955585)" | /rpt_family="L2" | misc_feature | /note="similar to EST AW553790 (NID:g7199213)" |
| misc_feature | 7955. .8046 | /note="similar to EST AW669129 (NID:g7525643)" | /rpt_family="L2" | misc_feature | /note="similar to EST AA475614 (NID:g2203465) vh22a10.r1" |
| misc_feature | 8131. .8208 | /note="similar to EST AW669129 (NID:g7525643)" | /rpt_family="L2" | misc_feature | /note="similar to EST A1012498 (NID:g3226330)" |
| misc_feature | 8519. .8602 | /note="similar to EST AW669129 (NID:g7525643)" | /rpt_family="L2" | misc_feature | /note="similar to EST AW918628 (NID:g8084409)" |
| misc_feature | 9308. .9528 | /note="similar to EST AA235790 (NID:g1860227) zs41g06.r1" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |
| misc_feature | 9341. .9659 | /note="similar to EST AA235984 (NID:g1860422) zs41q05.s1" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |
| misc_feature | 9836. .10096 | /note="similar to EST A1460700 (NID:g4112321) qx64h04.x1" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |
| misc_feature | 9838. .10417 | /note="similar to EST AA165701 (NID:g1741752) zo75g08.s1" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |
| misc_feature | 9838. .10241 | /note="similar to EST R45182 (NID:g8235536) yg40f08.s1" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |
| misc_feature | 9843. .10246 | /note="similar to EST AA563902 (NID:g2335541) nk20c04.s1" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |
| misc_feature | 9843. .10233 | /note="similar to EST A1198559 (NID:g37511165) qf49h08.x1" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |
| misc_feature | 9843. .10226 | /note="similar to EST R41923 (NID:g817622) yg04c10.s1" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |
| misc_feature | 9843. .10197 | /note="similar to EST A1418905 (NID:g4264836) tf43d02.x1" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |
| misc_feature | 9843. .10156 | /note="similar to EST A1858506 (NID:g5512122) wl39d09.x1" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |
| misc_feature | 9843. .10032 | /note="similar to EST A1989654 (NID:g5836535) ws26g11.x1" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |
| misc_feature | 9843. .9980 | /note="similar to EST AW451283 (NID:g6992059)" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |
| misc_feature | 9844. .10155 | /note="similar to EST A1393218 (NID:g4222765) tg14c10.x1" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |
| misc_feature | 9844. .10093 | /note="similar to EST A1655464 (NID:g4739443) tt13a10.x1" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |
| misc_feature | 9844. .10093 | /note="similar to EST A1655464 (NID:g4739443) tt13a10.x1" | /rpt_family="L2" | misc_feature | /note="similar to EST B2147901 (NID:g8610625)" |

| | | |
|--------------|--------------|---|
| misc_feature | 9845..10246 | /note="similar to EST A1655465 (NID:g4739444) ttl3a11.x1" |
| misc_feature | 9845..10236 | /note="similar to EST AA935178 (NID:g3092335) oh33d03.s1" |
| misc_feature | 9845..10236 | /note="similar to EST A1012498 (NID:g3226330)" |
| misc_feature | 9845..10236 | /note="similar to EST AW918628 (NID:g8084409)" |
| misc_feature | 9846..10236 | /note="similar to EST AW553790 (NID:g7199213)" |
| misc_feature | 9873..10236 | /note="similar to EST AA475614 (NID:g2203465) vh22a10.r1" |
| misc_feature | 9968..10224 | /note="similar to EST AA365758 (NID:g2018306)" |
| misc_feature | 9989..10498 | /note="similar to EST D78947 (NID:g1180820)" |
| misc_feature | 10054..10440 | /note="similar to EST AA302774 (NID:g1955159)" |
| misc_feature | 10058..10199 | /note="similar to EST AW451283 (NID:g6992059)" |
| misc_feature | 10060..10202 | /note="similar to EST AW205200 (NID:g6504681)" |
| misc_feature | 10218..10598 | /note="similar to EST AA984525 (NID:g3163050) am88c02.s1" |
| misc_feature | 10228..10764 | /note="similar to EST C18330 (NID:g1579932)" |
| misc_feature | 10230..11011 | /note="similar to EST W28969 (NID:g1308917)" |
| misc_feature | 10232..10456 | /note="similar to EST W67340 (NID:g1376209) zd43d07.r1" |
| misc_feature | 10265..10508 | /note="similar to EST AW553790 (NID:g7199213)" |
| misc_feature | 10265..10404 | /note="similar to EST AA475614 (NID:g2203465) vh22a10.r1" |
| misc_feature | 10265..10399 | /note="similar to EST A1012498 (NID:g3226330)" |
| misc_feature | 10265..10331 | /note="similar to EST AW918628 (NID:g8084409)" |
| misc_feature | 10309..10817 | /note="similar to EST BE147901 (NID:g8610625)" |

Query Match 97.1%; Score 33; DB 9; Length 135033;
Best Local Similarity 97.1%; Pred. No. 0.0075;
Matches 33; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGCTCCGGAGGACGCTGCCATGCCCAACT 34
|||||
Db 46065 AGGCTCCGGAGGACGCTGCCATGCCCAACT 46098

RESULT 14
AC024009
LOCUS AC024009 182662 bp UNA linear HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 2 clone RP11-158J14 map 2, WORKING DRAFT
SEQUENCE, 30 unordered pieces.
AC024009
VERSION AC024009.2 GI:7210067
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 182662)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 2, clone RP11-158J14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 182662)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhvalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepl,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeRellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenesor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galaguru,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McMan,P., McGuck,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,I.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

Direct Submission
Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 182662)
Hirren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larcocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,K., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,I.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (24-AUG-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 8, 2000 this sequence version replaced gi:7008914.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: 17216
Center clone name: 158_J_14
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 162604 bases at least Q40
Consensus quality: 171836 bases at least Q30
Consensus quality: 176600 bases at least Q20
Insert size: 180000; agarose-ff
Insert size: 179762; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-ff
Quality coverage: 3.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs, the true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

* be preserved.
1 37: contig of 37 bp in length
38 137: gap of 100 bp
138 1642: contig of 1505 bp in length
1643 1742: gap of 100 bp
3192: contig of 1450 bp in length
3193 3292: gap of 100 bp
3293 4982: contig of 1690 bp in length
4983 5082: gap of 100 bp
5083 5645: contig of 563 bp in length
5646 5745: gap of 100 bp
5746 7669: contig of 1924 bp in length
7670 7769: gap of 100 bp
7770 9530: contig of 1761 bp in length
9531 9640: gap of 100 bp
9641 11783: contig of 2153 bp in length
11784 11883: gap of 100 bp
11884 14433: contig of 2550 bp in length
14434 14533: gap of 100 bp
14534 18392: contig of 3859 bp in length
18393 18492: gap of 100 bp
18493 23382: contig of 4890 bp in length
23383 23482: gap of 100 bp
23483 27364: contig of 3882 bp in length
27365 27464: gap of 100 bp
27465 32120: contig of 4656 bp in length
32121 32220: gap of 100 bp
32221 36450: contig of 4230 bp in length
36451 36550: gap of 100 bp
36551 41108: contig of 4558 bp in length
41109 41208: gap of 100 bp
41209 45527: contig of 4319 bp in length
45528 45627: gap of 100 bp
45628 50262: contig of 4645 bp in length
50263 50362: gap of 100 bp
50363 55362: contig of 5000 bp in length
55363 55462: gap of 100 bp
55463 60928: contig of 5466 bp in length
60929 61028: gap of 100 bp
61029 66694: contig of 5666 bp in length
66695 66794: gap of 100 bp
66795 73126: contig of 6342 bp in length
73127 73226: gap of 100 bp
73227 79690: contig of 6464 bp in length
79691 79790: gap of 100 bp
79791 87490: contig of 7700 bp in length
87491 87590: gap of 100 bp
87591 95987: contig of 8397 bp in length
95988 96087: gap of 100 bp
96088 107334: contig of 11247 bp in length
107335 107434: gap of 100 bp
107435 117711: contig of 10277 bp in length
117712 117811: gap of 100 bp
117812 128059: contig of 10248 bp in length
128060 128159: gap of 100 bp
128160 143948: contig of 15789 bp in length
143949 144048: gap of 100 bp
144049 157783: contig of 13735 bp in length
157784 157883: gap of 100 bp
157884 182662: contig of 24779 bp in length.
----- Location/Qualifiers
1..182662
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-158J14"
1..37
/note="assembly_fragment
clone_end:T7
vector_side:right"
138..1642

FEATURES
Source
misc_feature
misc_feature


```

misc_feature /note="assembly_fragment"
1743..3192
/note="assembly_fragment"
3293..4982
/note="assembly_fragment"
5083..5645
/note="assembly_fragment
clone_end:SP6
vector_side:right"
5746..7669
/note="assembly_fragment"
7770..9530
/note="assembly_fragment"
9631..11783
/note="assembly_fragment"
11884..14433
/note="assembly_fragment"
14534..18392
/note="assembly_fragment"
18493..23382
/note="assembly_fragment"
23483..27364
/note="assembly_fragment"
27465..32120
/note="assembly_fragment"
32221..36450
/note="assembly_fragment"
36551..41108
/note="assembly_fragment"
41209..45527
/note="assembly_fragment"
45628..50626

```

Query Match 97.1%; Score 33; DB 2; Length 182662;

Best local Similarity 97.1%; Pred. No. 0.0074;

Matches 33; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 AGGCTCGGAGAACGNCCTGCCATGCCCACT 34
|||||
Db 90614 AGGCTCGGAGAACGNCCTGCCATGCCCACT 90647
|||||

```

```

RESULT 15
AC040991/c
LOCUS AC040991 186883 bp DNA linear HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 2 clone RP11-123E11 map 2, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
ACCESSION AC040991
VERSION AC040991.2 GI:8014664
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 186883)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
TITLE Homo sapiens chromosome 2, clone RP11-123E11
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 186883)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Bozulavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galaan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

```

O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 186883)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Bozulavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galaan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

TITLE

JOURNAL

COMMENT

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 22, 2000 this sequence version replaced gi:7534193.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/KM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9492

Center clone name: 123_E_11

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 175185 bases at least Q40

Consensus quality: 181490 bases at least Q30

Consensus quality: 183848 bases at least Q20

Insert size: 190000; agarose-fp

Insert size: 185283; sum-of-contigs

Quality coverage: 4.3 in Q20 bases; agarose-fp

Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1462: contig of 1462 bp in length

* 1463 152: gap of 100 bp

* 1563 4486: contig of 2924 bp in length

* 4487 4580: gap of 100 bp

* 4587 9251: contig of 4665 bp in length

* 9252 9351: gap of 100 bp

```
* 9352 12768: contig of 3417 bp in length
* 12769 12868: gap of 100 bp
* 12869 17380: contig of 4512 bp in length
* 17381 17480: gap of 100 bp
* 17481 22233: contig of 4753 bp in length
* 22234 22333: gap of 100 bp
* 22334 25600: contig of 3267 bp in length
* 25601 25700: gap of 100 bp
* 25701 31826: contig of 6126 bp in length
* 31827 31926: gap of 100 bp
* 31927 41356: contig of 9430 bp in length
* 41357 41456: gap of 100 bp
* 41457 53195: contig of 11739 bp in length
* 53196 53295: gap of 100 bp
* 53296 64559: contig of 11264 bp in length
* 64560 64659: gap of 100 bp
* 64660 79824: contig of 15165 bp in length
* 79825 79924: gap of 100 bp
* 79925 99909: contig of 19985 bp in length
* 99910 100009: gap of 100 bp
* 100010 121977: contig of 21968 bp in length
* 121978 122077: gap of 100 bp
* 122078 143182: contig of 21105 bp in length
* 143183 143282: gap of 100 bp
* 143283 164587: contig of 21305 bp in length
* 164588 164687: gap of 100 bp
* 164688 186883: contig of 22196 bp in length.
FEATURES
  source
    1..186883
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="2"
      /map="2"
      /clone="RP11-123E11"
      /clone_lib="RP11-11 Human Maie BAC"
      1..1462
        /note="assembly_fragment"
      1563..4486
        /note="assembly_fragment"
      4587..9251
        /note="assembly_fragment"
      9352..12768
        /note="assembly_fragment"
      12869..17380
        /note="assembly_fragment"
      clone_end:T7
      vector_side:left"
      17481..22233
        /note="assembly_fragment"
      22334..25600
        /note="assembly_fragment"
      clone_end:SP6
      vector_side:right"
      25701..31826
        /note="assembly_fragment"
      31927..41356
        /note="assembly_fragment"
      41457..53195
        /note="assembly_fragment"
      53296..64559
        /note="assembly_fragment"
      64660..79824
        /note="assembly_fragment"
      79925..99909
        /note="assembly_fragment"
      100010..121977
        /note="assembly_fragment"
      122078..143182
        /note="assembly_fragment"
      143283..164587
        /note="assembly_fragment"
      164688..186883
        /note="assembly_fragment"
```

```
BASE COUNT 42272 a 51993 c 50631 g 40377 t 1610 others
ORIGIN
Query Match 97.1%; Score 33; DB 2; Length 186883;
Best Local Similarity 97.1%; Pred. No. 0.0074;
Matches 33; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AGGCCTCCGGAGGACCGGNCCTGGCCATGTCACACT 34
|||||
Db 105563 AGGCCTCCGGAGGACCGGNCCTGGCCATGTCACACT 105550
|||||
```

Search completed: March 11, 2003, 09:15:02
Job time : 8549.13 secs